Lustre as a Data Acquisition file system at Diamond Light Source

Greg Matthews, EOFS Lustre Workshop presentation.

26 September 2011

diamond

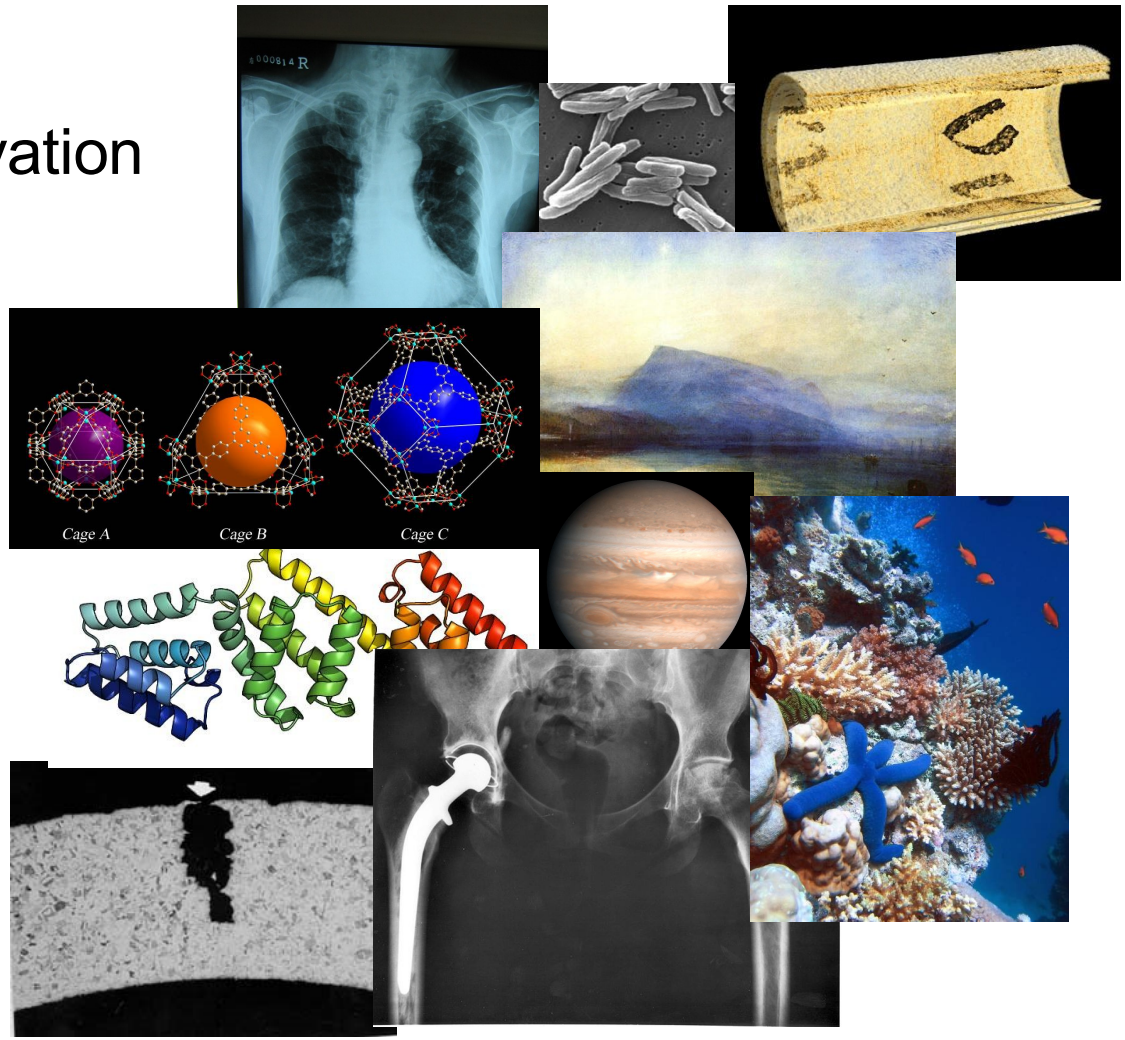
What is Diamond?

- The UK's national synchrotron facility.
- Extreme intensity light
- 562m circumference
- 3GeV
- 20 operational beamlines (more coming on stream)
- Variety of material science

diamond

Examples of science done at DLS

- Archeological preservation
- Bioscience research
- Climate change
- Nanotechnology
- “Green” technologies
- Extreme conditions
- Medical science
- Material corrosion



diamond

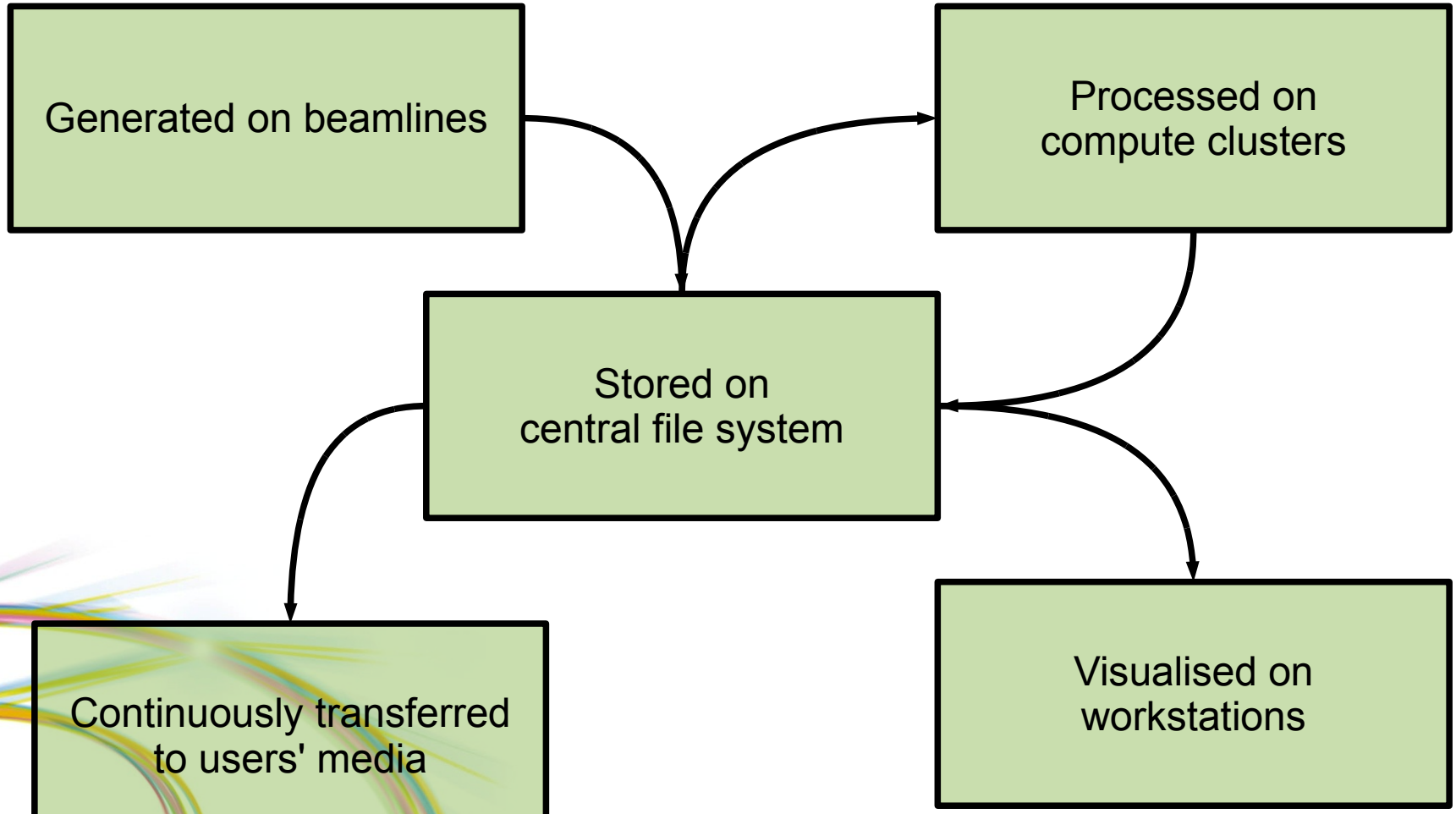
Science Computing

Science computing at Diamond includes:

- Data acquisition: beamline detectors
- Data storage: dependent on data rate
- Data processing: HPC
- Data transfer: provide a copy for users

diamond

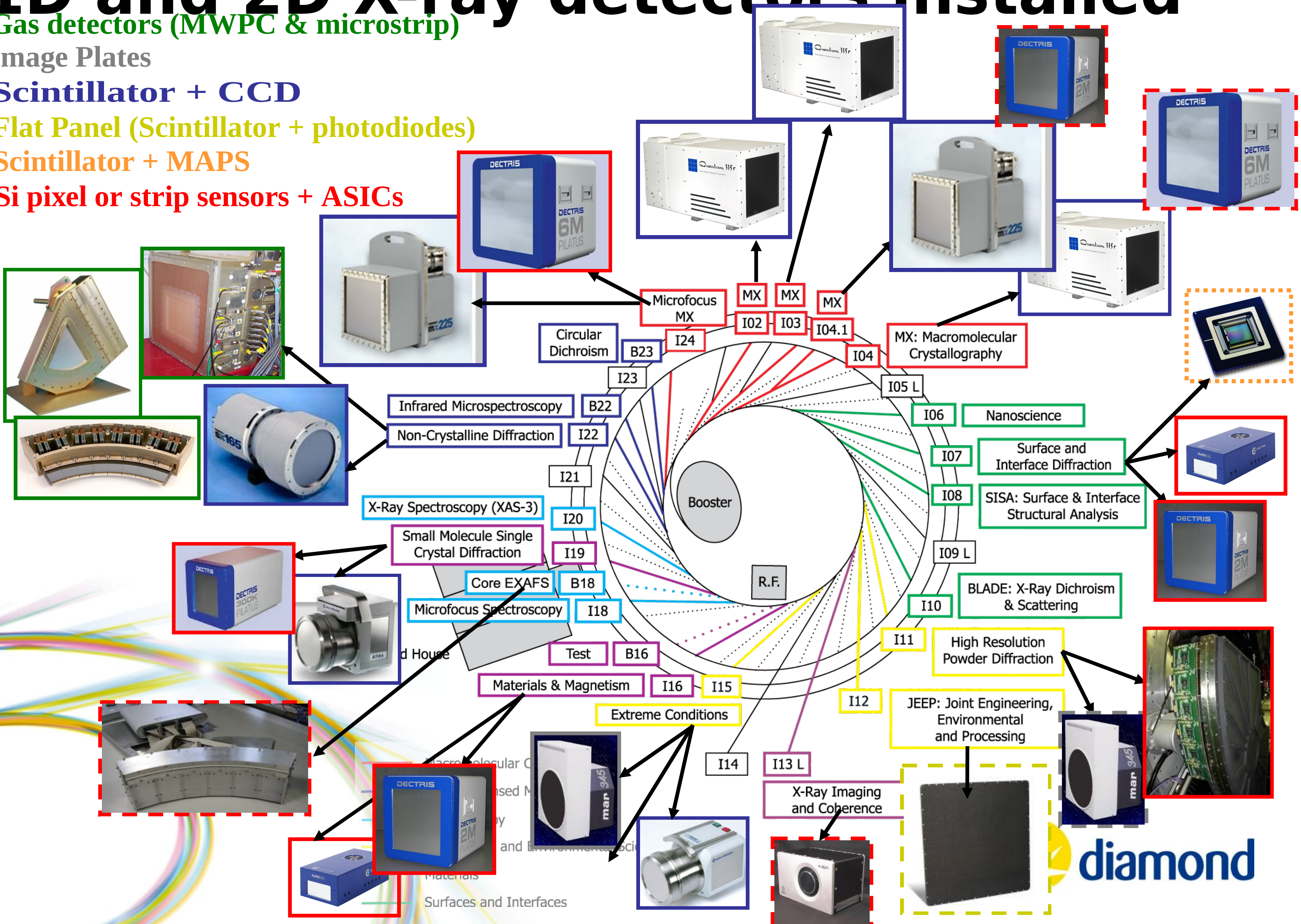
Data Flow



diamond

1D and 2D X-ray detectors installed

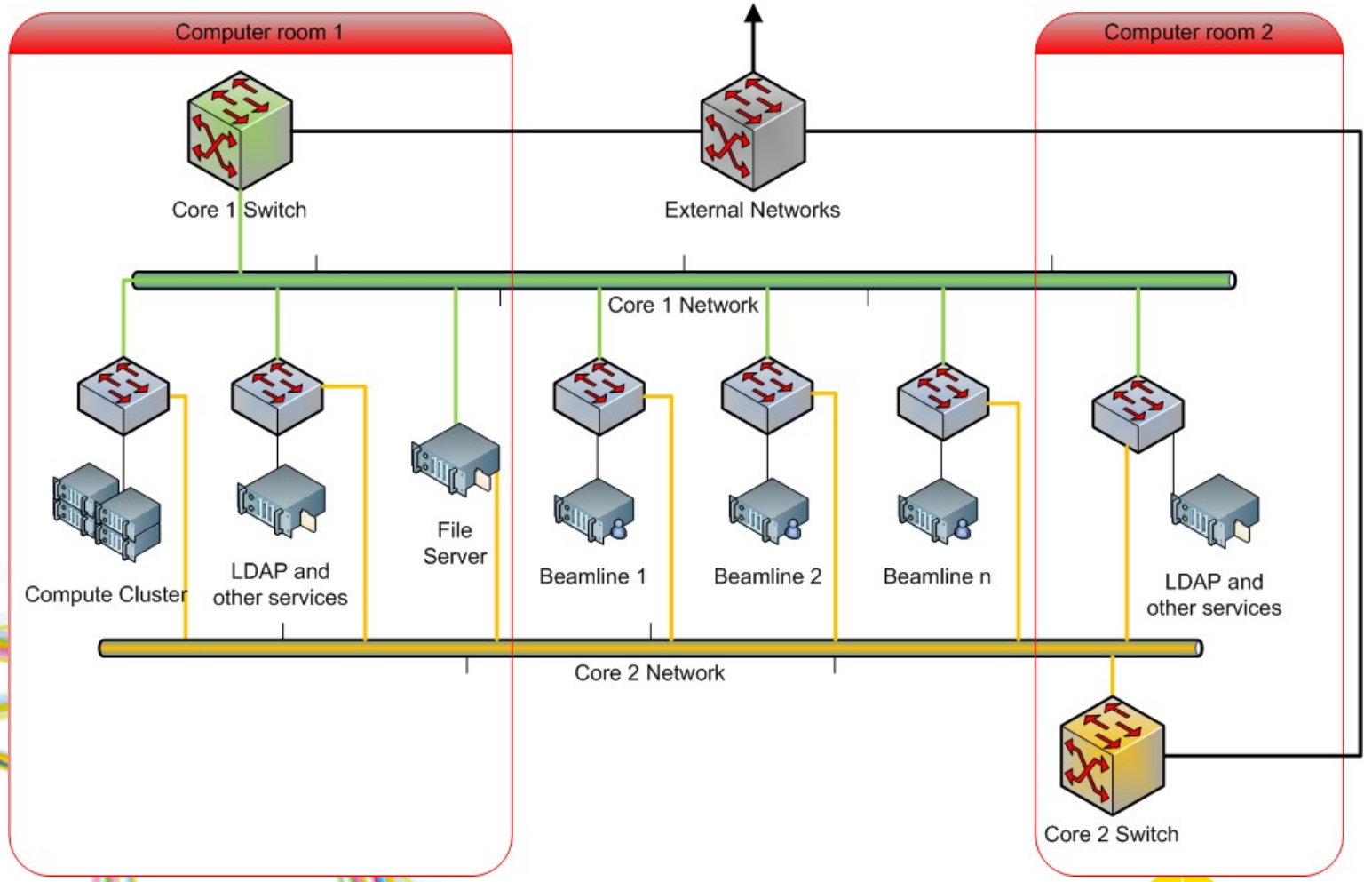
Gas detectors (MWPC & microstrip)
Image Plates
Scintillator + CCD
Flat Panel (Scintillator + photodiodes)
Scintillator + MAPS
Si pixel or strip sensors + ASICs



Data generation

diamond

Network Topology



diamond

Lustre storage #1

- 400TB raw (~300TB usable) > 60% full
- DDN S2A 9900 for OSTs, MD3000 for MDT
- PE2970 (Opteron) for OSS and MDS
- 6 OSS in fail-over pairs
- 2 MDS active/passive fail-over pair
- 10GbE to core networks
- Lustre 1.8.3-ddn3.3
- Achieved ~3.5GB/s write speed

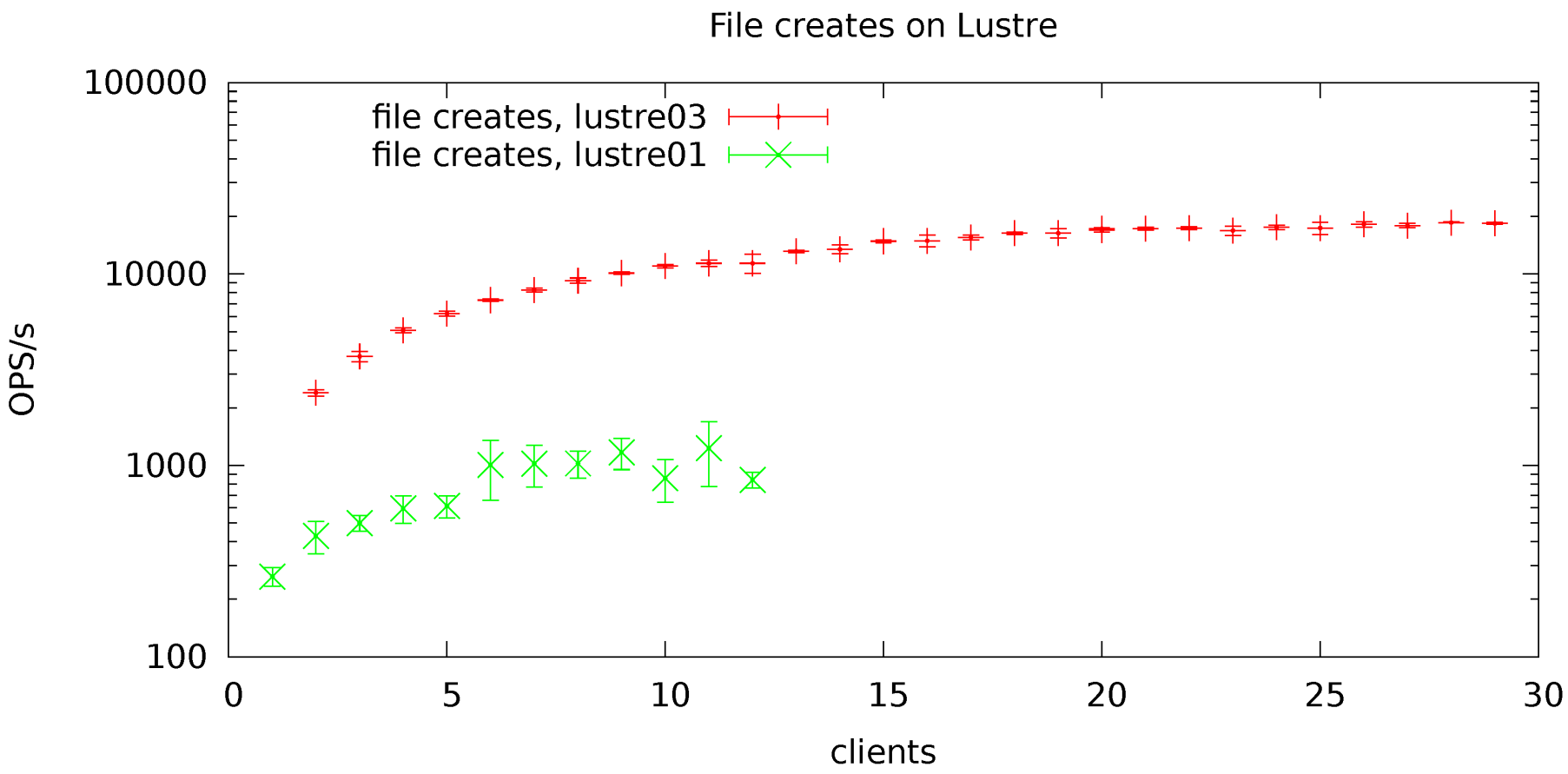
diamond

Lustre storage #2

- 600TB raw (~400TB usable)
- DDN SFA 10k for OSTs, EFI 3015 for MDT
- PE R610 for OSS and MDS
- 4 OSS in fail-over pairs
- 2x 10GbE channel bonded to core networks
- Lustre 1.8.4-ddn3.1
- ~5.5GB/s write speed
- Higher meta-data rate

diamond

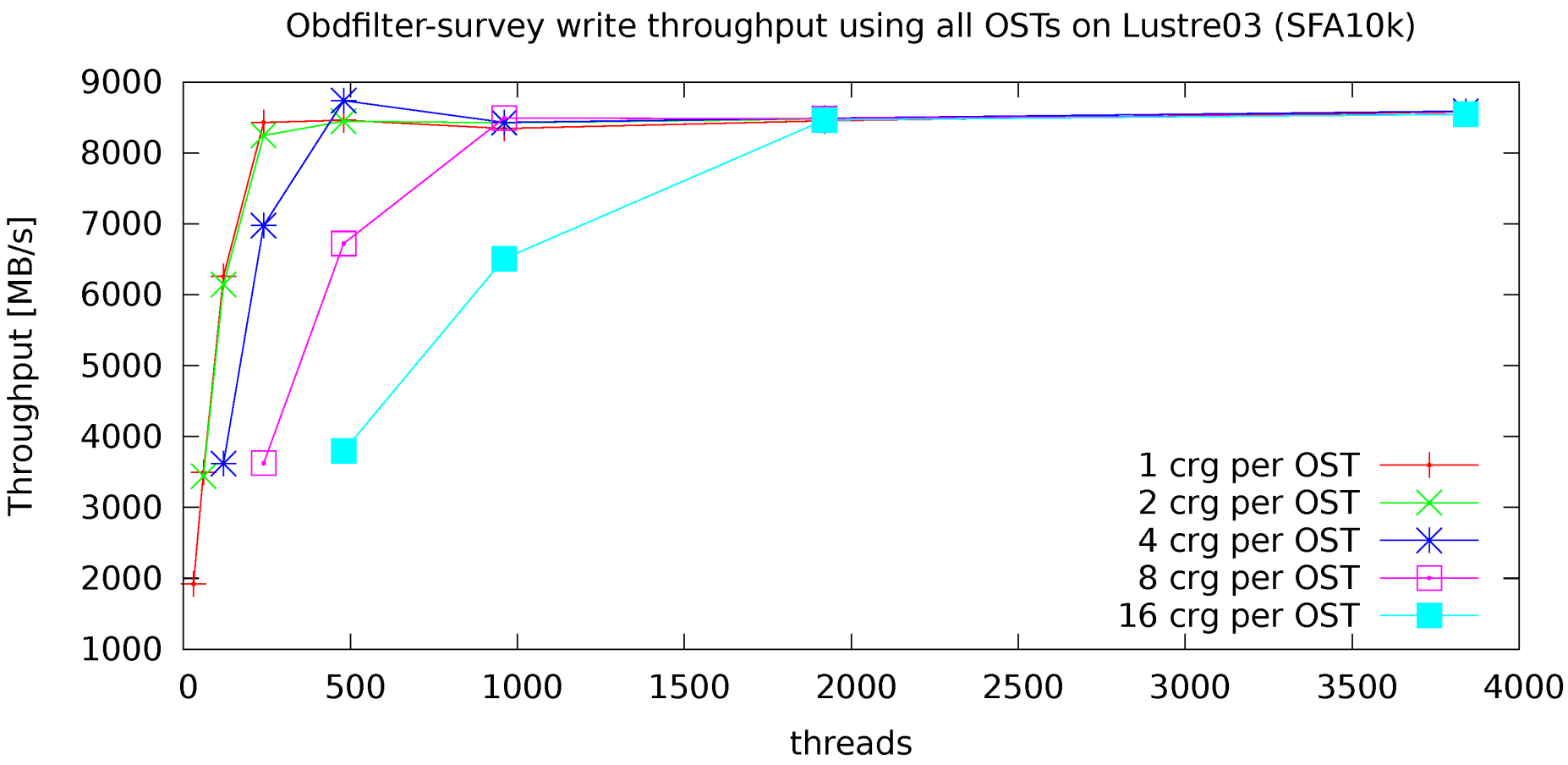
mdtest results



mdtest -I 10 -z 5 -b 5 -i 5 -u -d ${TESTDIR} creating 39060 objects per client

diamond

obdfilter-survey



diamond

The Lustre Experience

- Good:
  - Scalability
  - High throughput to/from Lustre clients
  - Suprisingly robust!
  - Able to use it on high data rate detectors (running non-DLS Linux OS)
  - Good access from compute clusters

diamond

The Lustre Experience

- Bad:
  - The (in)famous `ls -l`
  - Many small files
  - Small (4k) reads of large files
  - CIFS export
  - NFS export (LBUG)

diamond

Interest in development

- NFS export related LBUG
  http://jira.whamcloud.com/browse/LU-534
- lmt:
  - We are finding lmt to be a very useful tool but there are issues around using it with the MDT data.
- Windows client:
  - Detector controller PCs

We are interested to hear how we can help with development.

diamond

Has anyone here...?

- … heard of PanData? http://www.pan-data.eu/PaNdataODI
  - European money
    - “To develop a scalable data processing framework combining parallel filesystems with a parallelized standard data format (pNexus pHDF5) to permit applications to make most efficient use of dedicated multi-core environments and to permit simultaneous ingest of data from various sources, while maintaining the possibility for real-time data processing.”
  - Lustre is the obvious candidate
- …thought they'd quite like a Windows client?

Can EOFS help?

diamond

Thank You

Frederik Ferner who contributed many of the slides
Science Computing Team at Diamond: Nick Rees, Frederik Ferner and Tina Friedrich

diamond